A state of the art review on various malware detection and analysis in web security

Rajkumar E.V.[1] Aravindharamanan. S[2]
[1]PG Scholar
[1,2]Department of Information Technology
[1,2]Veltech Engineering College

Abstract— Securing the web is possibly a huge challenge that the modern era of computers have seen. Day by day the threat levels increase thereby making the network vulnerable to attacks. Many innovative strategies are brought into the field of cyber security to protect websites from attacks. But still malware has remained a serious cause of concern to web developers and server administrators. Alleviating this problem completely is a growing area of research. This paper is aimed at reviewing many malware detection systems, analyzing them and also depicts various ways of detecting them.

Key words: Malware, Propagation, Detection, Honeypot, Obfuscation, Privacy

I. INTRODUCTION

Technology has become an element key for today's life style where both business and research worlds completely rely on the technology and its applications. However like the other side of the coin, these developments have also opened the doors for the hacking and attacking community, and within a few years the malware has become a major security threat, affecting computers and networks widely [1].

Initially, the hackers and attackers started invading others computers just for fun they did not have any serious intention to look for any great gains, until online commerce gained its popularity especially in banking, financial transactions etc, which made the hacker to get financial gains [2], this has motivated the attackers, to work more and more to keep the machines infected as longer as possible, to get more financial gains and more valued information and data [2], consequently a big challenge has emerged in terms of protecting the information and business systems and a kind of arm races have started between security products and attackers community [3].

The malware historical timeline shows that it has a lot of changes and phases since it has been discovered and detected in hosts and networks , starting from virus which is a self-replicating malware but not self-transporting [2], moving to worm, which is a self-replicating and self-transporting [4], and going more for othermalware types and families . With the rapidly increasing complexity and interconnection of emerging information systems, the number of malware attacks is also increasing piercingly. While, there are a noticeable development in defense technologies and security techniques, there is also a similar development in sophisticated hacking techniques and appearance of new security vulnerabilities from day to day [5].

Many studies, surveys, experiments, brainstorming, statistical analysis and modeling methods have been done to gain deeper knowledge and valuable information about malware [15], because the attackers are continually developing their abilities, attacking skills and techniques. In order to make the tracking and detection processes difficult, and to pose new challenges to inspectors, all these studies and works are not sufficient enough to cover the rapid increase in malware evolution. Based on our understanding Virus Bulletin (1988) was the first dedicated Journal to study the malware [2], while, now there are a lot of Journals available that are dedicated to the security issues, especially malware issues. This paper has been presented to gain understanding about the various issues related to malware.

II. TYPES OF MALWARE

Recently, the number of information security threats caused by malware has rapidly increased, which leads to urgently studying the threats and accordingly categorizing them, to simplify the process of discovering and handling them, in order to detect them and find appropriate solutions. Malware has been categorized into seventeen different types [16], in this section we have listed and discussed the main and most common categories as follows:

A. Virus

Virus is a computer program that has the ability to harm and self-replicating in order to infect host; viruses are linked or attached to a software utility (e.g. PDF document). Launching the infected PDF document could then activate the virus, and a sequence of events may occur based on the function of the virus [4], [17].

B. Worm

Another kind of harmful programs is worm, which can replicate itself and invisibly transfer through networking. The effects of worms differ from viruses as the former need help from any file, to work and mainly its effect is on networking bandwidth or sending junk emails. One example of worms is Conficker [6], [17].

C. Spyware

this may occur, when users download free or trial software. In this kind, the users are observed by spies; hence their passwords, account numbers and every other personal detail become vulnerable [4].

D. Adware

this kind usually happens, while downloading free games or it is combined and embedded with advertisements, so when we watch advertisements this embedded code is installed to our PCs. This kind aims to observe the user's activities, when using networking [4].

E. Trojan

this kind gives power to remote hijackers, to use your system as they wish. They may get your passwords, observe your systems or damage the system files [4].

F. Botnet

this kind of malware controls your systems remotely and sends spam or spyware. Most of botnets are zombie and wait for command of the party who runs it, where there are two types of botnet such as, simple or hierarchical [18].

III. MALWARE PROPAGATION

Many studies and researches focused on studying the malware propagation in the digital world, communications and computer networks, some of the modeling and experimental procedures have been followed to study the effect of malware and the way it propagates in these fields, in addition to this, the studies cover some concepts and techniques related to malware detection. The malware propagation concept refers to the electronic method, by which, malware is transmitted to an information system, platform or device it seeks to infect for example the malware can propagate through PDF files and access the host unless the user disable the JavaScript in PDF reader [19].

A. Through Operating System

Malware is attacking the operating systems such as Mac, Android, Windows and Linux, but not in the same level and strength because some operating systems have more defense mechanisms which don't allow the malware to achieve its design purpose [20]–[22]. In the following lines some attacks followed by malware against OSs will be highlighted to show how the operating systems act accordingly. Every year a large number of new OSs malware with stronger propagation and strategies are created.

B. Through Wireless Networks

In [23], [24], the Authors have introduced the mobile and smartphone applications and some security issues related to wireless networks. In the above studies, the Bluetooth technology has been introduced in specific project named as Blue Bag that includes a covert attack and scanning device, which demonstrates how attackers can infect and reach a wide range of mobiles and devices running a Bluetooth Technology, they have found some weaknesses in Bluetooth technology, which may allow attackers to reach the devices. In [9] the authors have explained some specific attacks that can affect the wireless communication and Bluetooth such as:

1) BlueSnarf

It uses the (Object exchange) push service and the attacker can access without any authentication and recently in the upgraded version of this kind of attack, the attacker can get a full access including read and writes access.

2) Bluejacking

occurs by sending a short tricky text message into authenticated dialog, and the users will be using the access codes of the tricky message, which allows the attacker to take control of the device.

3) BlueBug

The attacker will be able to use phone services, which include incoming and outgoing calls, sending and receiving SMS, etc. all through accessing the cell phone.

Blue Bump: it goes through the weakness of Bluetooth in the way it handles link keys, and it can lead to getting the data or abusing the mobile services such as internet, WAP and GPRS.

4) Blue Smack:

It simply guides to service denial.

5) HeloMoto

It is a combination of BlueBug and BlueSnarf effect.

6) Blue Dump

The attacker will involve himself in the pairing process through Bluetooth after dumping the stored link key.

7) Car Whisperer

The default configuration of some devices makes the PIN code fixed for pairing and exchanging, which will make it easy for the attackers, to abuse the devices and take control of the devices accordingly once they get the PIN, which is not changeable.

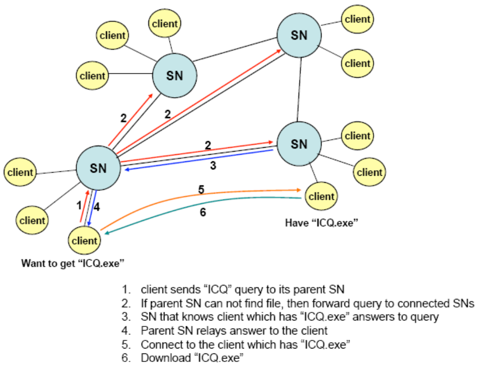
8) Blue Chop

The attacker will get the chance to disconnect and terminate the established connection, especially when the master of the connection is supporting multiple connections.

C. Through File Sharing

File sharing has become a very common application for Peer-to-Peer networking, which allows the users to share a huge number of digitally stored information, one of the most common file sharing networks is Kazaa, which has been developed in 2001, based on the Fast Track Protocol, Kazaa was subsequently underlicense as a legal music subscription service, but as of August 2012, the Kazaa website is not offering a music service anymore [6].

Having few number of defense mechanisms is the reason behind the vulnerability of the Peer-to-Peer file sharing networks to many security attacks ; according to this hundreds of viruses have used the P2P as a propagation vector , the authors have described how KaZaA works and shares files and explained the concept of supernode and indexing process for the hosts, where the connection between hosts is encrypted with a key exchanged at the beginning of the session , also they have discussed about Krawler (A KaZaA Crawler), which has two main components : the dispatcher, which maintains a list of super nodes and the fetcher, which is responsible for communicating with the Dispatcher , Updating process and Sending Queries. Fig. 1 below is an example of the search sequence in KaZaA.



1. client sends "ICQ" query to its parent SN
2. If parent SN can not find file, then forward query to connected SNs
3. SN that knows client which has "ICQ.exe" answers to query
4. Parent SN relays answer to the client
5. Connect to the client which has "ICQ.exe"
6. Download "ICQ.exe"

Fig. 1: Example of KaZaA Search Sequence

D. Through Social Networking

During the last few years, online social networks have become very popular and grew tremendously as they act as platform of real-world relationship; its popularity comes from the feature of virtual interaction techniques [8].

According to [7], OSN provides the users with many services such as, sharing photos, clips, files and applications in addition to chat and call services, and the last two years have proved that, the OSNs are not only websites for communication and fun, but can contribute in the process of changing the culture and lifestyle. On the other hand in terms of security, OSNs can be considered as a perfect environment for malware and security threats. Based on the studies and researches, the attacks and threats can be categorized into four wide categories:

(1) Privacy Breach Attacks: three primary parties interact with one another in an OSN: breaches from service providers, which represent the companies such as Facebook, Twitter and so on, breaches from other users and accounts owners, and breaches from third-party applications, which are involved in many stages of OSNs. On the other hand the threats related to the privacy issues can be classified as: browsing user activities, disclosing the user's identity, cyber-stalking, cyber-bullying, harassing, and slandering.

(2) Viral Marketing: this refers to the techniques of marketing, including OSNs and other technologies, in OSNs viral marketing can be considered as an unwanted and good environment for malware, one of the most common examples is the spam in OSNs, in addition to the process of phishing attacks, which is considered as social engineering technique [26].

(3) Network Structural Attacks: such as Sybil Attacks and here some defense mechanisms are provided such as, trusted certificates, resource testing and recurring costs.

(4) Malware Attacks: one of the most common attacks is the attack of a worm known as Koobface worm.

E. Through Virtualized Systems

Virtualization technique is quickly becoming a standard technique for business. The technology lets one computer or server run multiple operating systems, or multiple sessions of an operating system at the same time, which lets users run many applications and functions on a single computer or server, instead of running them on different machines as in the old technology.

The biggest challenge faced by organizations now is, how to secure the virtualized system, which are vulnerable to the same type of threats as real systems. Virtualized systems cannot always be secured by the same technique as real systems, because each virtualized system on the same machine may face different threats and need different security levels, and we need additional security techniques, to secure the channels between the virtualized systems on the same machine.

F. Through Email Communications

There are many ways to attack emails, which affects the sending emails (email backscatter) i.e. spam emails using viruses or worms. For that, we need to inform the sender about the real reasons for not receiving email from the other side. The attackers intercept the email, and delete the sender's address, therefore the email gets spammed and the receiving process fails, thus the sender receives a failing note message and he/she cannot determine the real reason of failure.

The email spam propagation can be analyzed by many factors such as the period of time between sending the email and sending back the failing report for the sender, another factor is the returned message which does not contain a real failing reason i.e. the system is down at this time please try later again [28].

IV. Malware Detection Techniques

Since malware has different types, behaviors and different level of risk, the same detection methods and mechanisms cannot be used in all cases. It is impractical to have just one security software to efficiently handle the malwares. Hence having different detection methods for different environments becomes unavoidable. This study had focused on the most common and powerful techniques such as honeypot, honeynet, virtualization (partial and full), sandboxing and behavior operation sets. A massive experiment had been done by Taiwan malware analysis net (TWMAN), it was based on virtualization concept and client-server model, the experiment added a great value to the field of malware detection since it was able to detect many malwares which were not detectable by normal detection methods, going forward, we can clearly see that the detection process needs more computer processing power and advance techniques to make sure that the nature and behavior of malware are clear and covered from all the angles and views.

A. Anomaly-Based

Anomaly-based detection looks for unexpected or abnormal behavior indicators, which indicate the presence of malware. In more detail, anomaly based detection creates a baseline of expected operation. After this baseline has been created, any different form of baseline is recognized as malware. We have identified that the anomaly based detection technique uses the previous knowledge of what is known as normal to find out what is malicious. A special type of anomaly based detection techniques is specification based detection.

B. Honeypots

The traditional methods for detecting and preventing malware, like using anti-virus can only detect the malware with the same features. In this method, security vendors build pattern files, which contain the features of malware that have been already collected and analyzed. However, it is not possible to detect malware with different features and characteristics, especially with increasing the variation in the ratio of malware [29].

C. Sandboxing

Previously, we had shown that the malware can exploit VME to propagate between VM hosts. In this section we will present how the VME can be used to detect malware and prevent its propagation.

There are many methods to prevent malware from detecting VME. In [27] [33], the authors have discussed two mainly useful methods to prevent the most popular VME detection techniques used by malicious attacker from detecting the VME and mitigate malware effect. After many experiments, a lot of undocumented configuration files were and the amazing result was that changing some parameters in this undocumented configuration files can prevent or control the behaviors that allow malware to detect VMware. For example Jerry.c can be prevented from being detected by VMware by setting VMX file parameters as in the following program snippet:

- Isolation.tools.getPtrLoc
- Ation.disable=”TRUE”
- Isolation.tools.setPtrLoc
- Ation.disable=”TRUE”
- Isolation.tools.setVersio
- n.disable=”TRUE”
- Isolation.tools.setPtrVersio
- n.disable=”TRUE”

D. Mathematical Models

In [10], differential equation and stochastic variance have been used to express the behavior of internet worm and modeling process, the model focuses on times of next infection (TNI), which is used to explain and clarify the variance. The paper contributed in this field by validating the infection times of the (TNI) with respect to oriented scanning model, based on the structure of Red Code, and to experimentally evaluate the variance using commonly used metrics for the process of detecting worms. Based on the results of the experiments, it shows that the level of variance is tremendously high; this variance should be taken in consideration.

V. Malware Analysis

Analyzing a malware is the inspection of the malware from its signature or behavior, to discover the attributes and functionalities of the malware; and to find out the source, target range, propagation approach and defense mechanisms of the malware. The result of these inspections helps increasing the security of the end users by providing better security through products like anti-viruses, intrusion detection systems and firewalls. Antivirus software usually maintains a virus signatures repository, which contains the binary patterns characteristic for the malicious codes. This software checks the files that are assumed to be infected for the existence of a virus signature. This detection method worked effectively until creator of malware started writing polymorphic and metamorphic code. These modifications of malware code enabled them to avoid detection by using encryption techniques, to prevent signature based detection. Security products and virus scanners look for the sequence of characteristic bytes (signature), to recognize the malicious code.

A. Malware Behavior

Behavior based detection techniques study and analyzes the behavior of suspected or known malicious code, such as destination and source addresses of this code, and the way in which, the code was attached. Behavior based detection technique differs from the other scanning techniques as it considers the action performed by the malware, rather than the binary pattern. The programs with different binary content but having same behavior are collected. These types of detection techniques help in detecting the malware, which keeps on generating new signature versions, because they will always use the recourses of the system in the same manner. The behavior detector collects the data, interprets the data, and then applies the matching algorithm [1], [35].

B. Malware Signature

Normal antivirus software look for signatures, which are a sequence of bytes in the malware code to state that if the program scanned, is malicious or not. Essentially, there are three types of malware: basic, polymorphic, and metamorphic malware. In basic malware, the malware developer changes the entry point of the program. Polymorphic viruses alter themselves, while leaving the original code unchanged. A polymorphic virus contains an encrypted malicious code beside the decryption part. This virus is enabled by a polymorphic engine, which is included in the body of the virus. The polymorphic engine generates new versions every time it is run; thus it is very difficult to detect this type of virus by signature based detection techniques. Metamorphic malware use advanced obfuscation techniques, to reprogram itself therefore the children and parent signatures are very different. It is not possible to detect this type of malware without disassemble the virus file [4], [38], [39].

C. Obfuscation and Normalization

It is a technique used by software developers and writers targeting to hide the details of their products so that the reverse engineers can’t find the correct code, it has been used as an advantage by the malware writers to achieve the same goal, obfuscation can be achieved by different operations and easily can make changes in the signature of malware in order to make the process of detecting the malware very difficult. Fig. 2 shows the obfuscation process [4], [40]. Given a program P and a transformation function T generates program P’ such that the following properties hold true:

- P’ is difficult to reverse engineer.
- P’ holds the functionality of P.
- P’ performs comparable to P

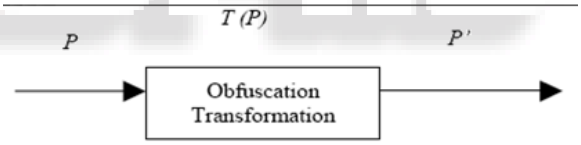


Fig. 2: The obfuscation process

The obfuscation techniques can be done in different methods, starting from inserting some (no operation) instructions and inserting (push-pop) x, which known as dead-code because nothing will be achieved and accomplished and inserting some instructions for branching unconditionally, moving to inserting process for some registers and substituting instructions, all of these methods will guide to obfuscate the code of the malware and make the process of detection difficult to malware scanners. Malware normalization can be identified as a process and mechanism to detect the obfuscated copies of malware and increasing the rate of catching the malware by the detector, the output of the normalization will be the original signature of the malware which has been obfuscated and accordingly the signature will be compared to the signatures to verify it, then it will be saved in the list of known signatures in order to decrease the time of scanning and detecting next times.

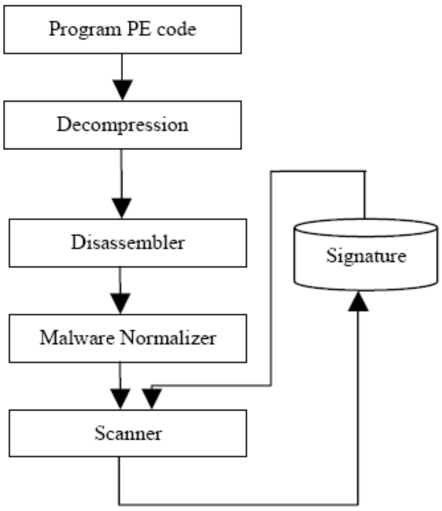


Fig. 3: Normalization process flow chart

VI. Conclusion And Future Work

The malware developer tries to write new techniques and strategies to hide the malicious code and infect the targets. On the other hand, the detectors analyze malware behaviors continuously and try to resist these techniques and strategies hence, we need to allow detection development techniques to lead malware updating through very well analytical process for malware activities and behaviors to fix any possible targeted threats. A new simulation must be designed to contain real system samples, to analyze the malware behaviors against these samples after elaborate malware updating. The objectives of this simulation are to avoid systems threats before being infected by real malware.

References

[1] L. Wu, R. Ping, L. Ke, and D. Hai-xin, "Behavior-Based Malware Analysis and Detection," in Complexity and Data Mining (IWCDM), 2011 First International Workshop on, 2011, pp. 39–42.
[2] R. Ford and W. H. Allen, "Malware Shall Greatly Increase...," Secur. Priv. IEEE, vol. 7, no. 6, pp. 69–71, 2009.
[3] J. R. Crandall, R. Ensafi, S. Forrest, J. Ladau, and B. Shebaro, "The ecology of Malware," in Proceedings of the 2008 workshop on New security paradigms, 2009, pp. 99–106.
[4] P. Vinod, R. Jaipur, V. Laxmi, and M. Gaur, "Survey on malware detection methods," in Proceedings of the 3rd Hackers' Workshop on Computer andInternet Security (IITKHACK'09), 2009, pp. 74–79.
[5] J.-H. Park, M. Kim, B.-N. Noh, and J. B. Joshi, "A Similarity based Technique for Detecting Malicious Executable files for Computer Forensics," in Information Reuse and Integration, 2006 IEEE International Conference on, 2006, pp. 188–193.
[6] S. Shin, J. Jung, and H. Balakrishnan, "Malware prevalence in the KaZaA file-sharing network," in Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, 2006, pp. 333–338.
[7] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security issues in online social networks," Internet Comput. IEEE, vol. 15, no. 4, pp. 56–63, 2011.
[8] S. Mohtasebi and A. Dehghantanha, "A Mitigation Approach to the Privacy and Malware Threats of Social Network Services," in Digital Information Processing and Communications, Springer, 2011, pp. 448–459.
[9] L. Carettoni, C. Merloni, and S. Zanero, "Studying bluetooth malware propagation: The bluebag project," IEEE Secur. Priv., vol. 5, no. 2, pp. 17–25, 2007.
[10] D. M. Nicol, "The impact of stochastic variance on worm propagation and detection," in Proceedings of the 4th ACM workshop on Recurring malcode, 2006, pp. 57–64.
[11] S. Zanero, "Wireless malware propagation: A reality check," Secur. Priv. IEEE, vol. 7, no. 5, pp. 70–74, 2009.
[12] C. Greamo and A. Ghosh, "Sandboxing and virtualization: Modern tools for combating malware," Secur. Priv. IEEE, vol. 9, no. 2, pp. 79–82, 2011.
[13] S. J. Vaughan-Nichols, "Virtualization sparks security concerns," Computer, vol. 41, no. 8, pp. 13–15, 2008.
[14] J. Van Randwyk, K. Chiang, L. Lloyd, and K. Vanderveen, "Farm: An automated malware analysis environment," in Security Technology, 2008. ICCST 2008. 42nd Annual IEEE International Carnahan Conference on, 2008, pp. 321–325.
[15] M. Apel, C. Bockermann, and M. Meier, "Measuring similarity of malware behavior," in Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on, 2009, pp. 891–898.
[16] M. F. Zolkipli and A. Jantan, "An approach for malware behavior identification and classification," in Computer Research and Development (ICCRD), 2011 3rd International Conference on, 2011, vol. 1, pp. 191–194.
[17] M. R. Rieback, P. N. Simpson, B. Crispo, and A. S. Tanenbaum, "RFID malware: Design principles and examples," Pervasive Mob. Comput., vol. 2, no. 4, pp. 405–426, 2006.
[18] B. Stone-Gross, M. Cova, B. Gilbert, R. Kemmerer, C. Kruegel, and G. Vigna, "Analysis of a botnet takeover," Secur. Priv. IEEE, vol. 9, no. 1, pp. 64–72, 2011.
[19] D. Stevens, "Malicious PDF documents explained," Secur. Priv. IEEE, vol. 9, no. 1, pp. 80–82, 2011.
[20] J. Boutet, "Malicious Android Applications: Risks and Exploitation," Inst. InfoSec Read. Room, vol. 2, 2010.
[21] A. J. O'Donnell, "When malware attacks (anything but windows)," Secur. Priv. IEEE, vol. 6, no. 3, pp. 68–70, 2008.
[22] J. Yonts, "Mac OS X Malware Analysis," Inst. InfoSec Read. Room, vol. 2, 2009.